

ViRBase: a resource for virus–host ncRNA-associated interactions

Yanhui Li^{1,2,†}, Changliang Wang^{1,†}, Zhengqiang Miao^{1,†}, Xiaoman Bi^{1,†}, Deng Wu^{1,†}, Nana Jin^{1,†}, Liqiang Wang¹, Hao Wu¹, Kun Qian¹, Chunhua Li¹, Ting Zhang¹, Chunrui Zhang¹, Ying Yi¹, Hongyan Lai¹, Yongfei Hu¹, Lixin Cheng^{3,4}, Kwong-Sak Leung³, Xiaobo Li⁵, Fengmin Zhang⁶, Kongning Li^{1,*}, Xia Li^{1,*} and Dong Wang^{1,*}

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, ²Institute of Cardiovascular Sciences and Key Laboratory of Molecular Cardiovascular Sciences, Peking University Health Science Center, Beijing, China, ³Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China, ⁴State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong, China, ⁵Department of Pathology, Harbin Medical University, Harbin, China and ⁶Department of Microbiology, Harbin Medical University, Harbin, China

Received August 01, 2014; Revised September 12, 2014; Accepted September 18, 2014

ABSTRACT

Increasing evidence reveals that diverse non-coding RNAs (ncRNAs) play critically important roles in viral infection. Viruses can use diverse ncRNAs to manipulate both cellular and viral gene expression to establish a host environment conducive to the completion of the viral life cycle. Many host cellular ncRNAs can also directly or indirectly influence viral replication and even target virus genomes. ViRBase (<http://www.rna-society.org/virbase>) aims to provide the scientific community with a resource for efficient browsing and visualization of virus–host ncRNA-associated interactions and interaction networks in viral infection. The current version of ViRBase documents more than 12 000 viral and cellular ncRNA-associated virus–virus, virus–host, host–virus and host–host interactions involving more than 460 non-redundant ncRNAs and 4400 protein-coding genes from between more than 60 viruses and 20 hosts. Users can query, browse and manipulate these virus–host ncRNA-associated interactions. ViRBase will be of help in uncovering the generic organizing principles of cellular virus–host ncRNA-associated interaction networks in viral infection.

INTRODUCTION

Non-coding RNAs (ncRNAs) are important functional RNA molecules (1–3). Recent advances in virology reveal that ncRNAs play critical roles during almost every process of viral infection, such as regulation of virus growth, replication and cell death (2,4,5). Accumulated evidence shows that viral ncRNAs can be used to manipulate both cellular and viral gene expression to establish a host environment conducive to the completion of the viral life cycle (2,6). Similarly, many cellular ncRNAs can also directly or indirectly influence viral replication and even target viral genomes (2,7,8). Interaction of the viral and cellular ncRNAs with their viral and cellular targets forms a complex functional interaction network. Elucidation of such complex interactions is fundamental for understanding viral infection and developing new antiviral therapies.

To facilitate related research in virology, we developed a virus–host ncRNA-associated interaction database (ViRBase; <http://www.rna-society.org/virbase>), aimed at collecting viral and cellular ncRNA-associated virus–virus, virus–host, host–virus and host–host interactions in viral infection by manually curating the literature. The current version of ViRBase documents over 12 000 viral and cellular ncRNA-associated virus–virus, virus–host, host–virus and host–host interactions involving more than 60 different viruses and 20 hosts. Most medically relevant viruses are contained within the database, including Epstein–Barr virus, human immunodeficiency virus 1 (HIV-1), hepatitis B virus (HBV), hepatitis C virus (HCV), herpes viruses, influenza A virus H1N1 and H3N2, Kaposi's sarcoma-

*To whom correspondence should be addressed. Tel: +86 451 8661 5922; Fax: +86 451 8661 5922; Email: wangdong@ems.hrbmu.edu.cn
Correspondence may also be addressed to Xia Li. Tel: +86 451 8661 5922; Fax: +86 451 8661 5922; Email: lixia@hrbmu.edu.cn
Correspondence may also be addressed to Kongning Li. Tel: +86 451 8661 5922; Fax: +86 451 8661 5922; Email: kongningli@hotmail.com

[†]The authors wish it to be known that, in their opinion, the first six authors should be regarded as Joint First Authors.

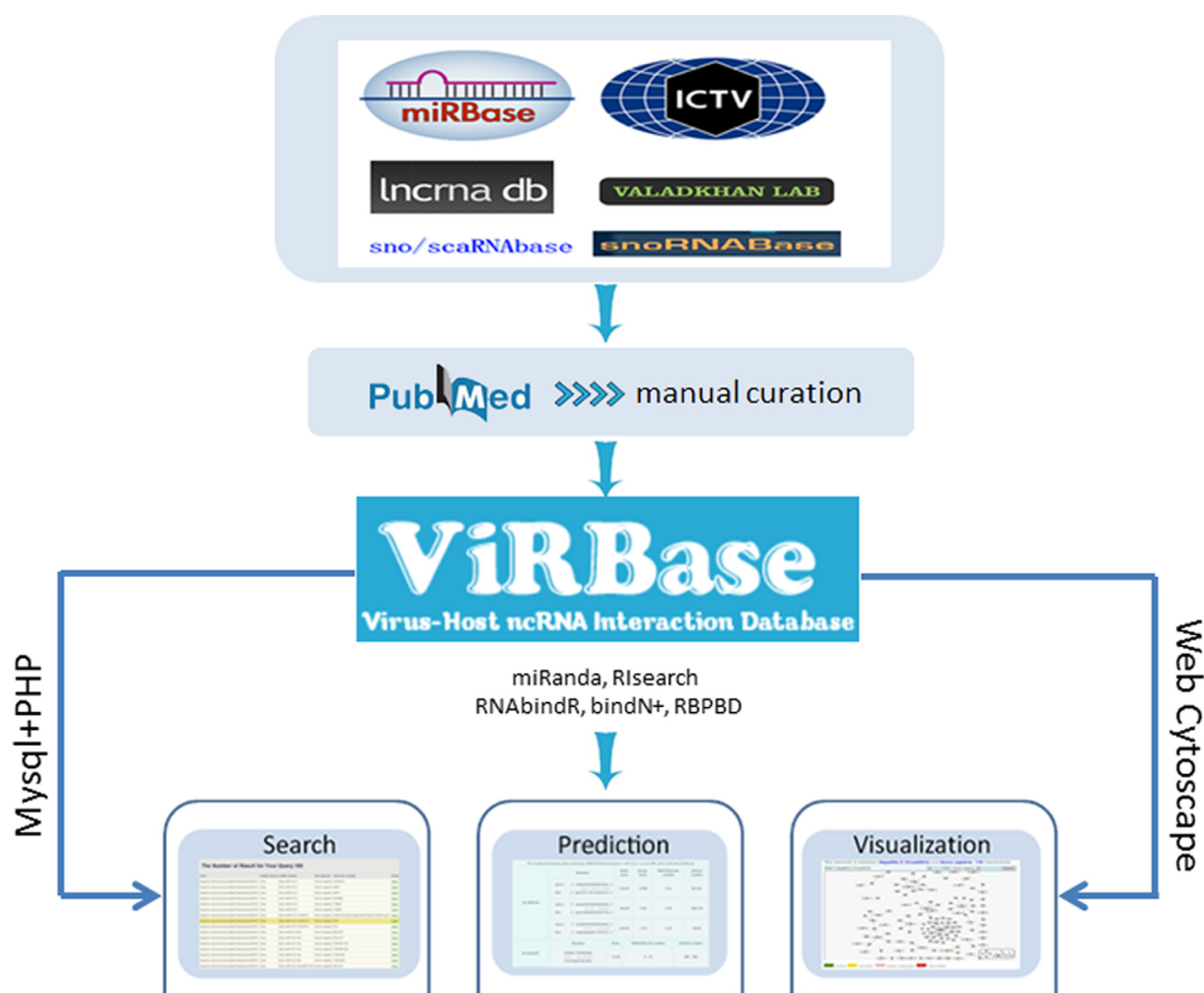


Figure 1. The overview of the ViRBase database.

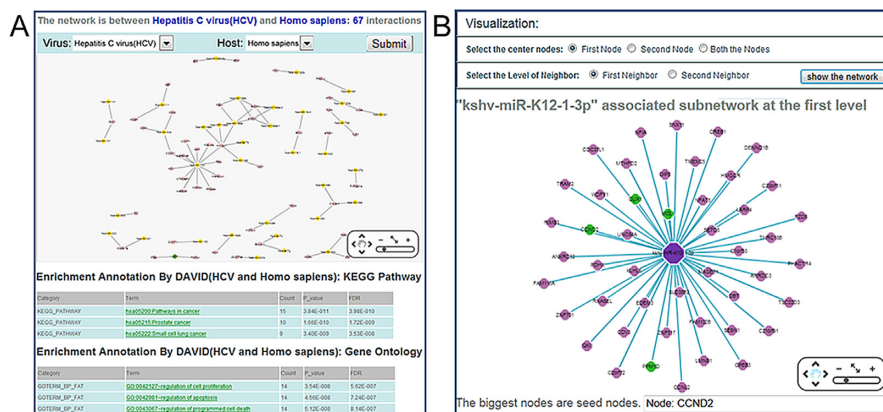


Figure 2. Representative screenshots of the Visualization and Network pages. (A) The Visualization page: representing the virus–host ncRNA-associated interaction network for a specific virus and host. At the bottom, the potential Gene Ontology functions and KEGG pathways by performing enrichment analysis are provided. (B) The Network page: representing the specific virus–host ncRNA-associated interaction sub-network.

Predicted Binding Sites



Figure 3. Representative screenshots of the Binding pages. The Binding page represents the predicted binding sites and scores between ncRNAs and their corresponding interactor by various of the predictor tools.

associated herpesvirus, papilloma viruses and simian virus 40. Hence, ViRBase provides a global view of virus–host ncRNA-associated interaction networks in viral infection. Researchers can follow these interactions to explore how the virus–host ncRNA-associated interaction network is organized. The whole data set can be easily queried and downloaded through the webpage, and visualization tools for interactively browsing and analyzing the data set are provided. In addition, ViRBase also allows researchers to submit new virus–host ncRNA-associated interactions.

DATA SOURCES AND IMPLEMENTATION

In order to collect all available ncRNAs, we have integrated information on three major types of ncRNAs: long non-coding RNA (lncRNA) symbols from lncRNAdb (9) and the functional lncRNA database (www.valadkhanlab.org/database/), microRNA symbols from the mirBase database (10) and small nucleolar RNA (snoRNA) symbols from sno/scaRNAbase (11) and snoRNA-LBME-db (12). We also used the ncRNA category names for other ncRNAs, such as transfer RNA and small nuclear RNA. The list of virus names and abbreviations was collected according to international committee on taxonomy of viruses database (ICTVdb) (13,14). We have written a simple script to screen all abstracts and articles in the PubMed database using the following keyword combinations: (each ncRNA symbol or ncRNA category name) and/or (each virus) and/or (interaction or binding, etc.). The relevant hits were further inspected manually.

The ViRBase database is implemented in the HTML and PHP languages with MySQL server. The interface component consists of webpages designed and implemented in

HTML/CSS in a Microsoft Windows environment. It has been tested in the Google Chrome, Firefox and Internet Explorer web browsers.

CONTENT OF THE DATABASE

All interaction information was obtained from articles published in the PubMed database before August 2014. In the current version, ViRBase documents 12 247 ncRNA-associated virus–virus, virus–host, host–virus and host–host interactions in viral infection (including 11 816 experimental interactions and 431 predicted interactions), involving 462 non-redundant ncRNAs and 4463 protein-coding genes from 61 viruses and 24 hosts. Each entry contains detailed information on a virus–host ncRNA-associated interaction, including virus name, host species, ncRNA symbol, target symbol, validation method, Pubmed ID and detailed description (Figure 1).

ViRBase also provides five options on the ‘Help’ page to provide instructions for using the database. These include ‘Tutorial’ (procedure and illustrations of the database), ‘Source’ (sources of ncRNAs and gene information, and tools used in the database), ‘Parameter’ (details the parameters for tools used for binding site prediction), ‘Statistics’ (detailed statistical tables) and ‘Error Report’. In the ‘Download & API’ page, users can download all interaction data in Microsoft Excel and TXT format by selecting ‘ALL Virus–Host Interactions’ or access the application programming interface (API) using scripts. In the ‘Submit’ page, ViRBase invites users to submit novel virus–host ncRNA-associated interactions.

DATA QUERYING, SEARCHING AND BROWSING

ViRBase provides an interface for convenient retrieval of all interactions. Users can search each virus–host ncRNA-associated interaction through four paths (Supplementary Figure S1), including ‘By keyword’ (search by any key information with support for fuzzy search), ‘By Virus and Host Species’ (select virus in ICTV taxon or host of interest), ‘By RNA/Protein Symbol’ (select RNA/Protein symbols based on the categories with multiple selection supported) and ‘By Validated Methods’ (select or input validated methods of interest; fuzzy search and multiple selections are supported). In the ‘By Validated Methods’ path, we also distinguish between experimentally proven and predicted interactions. Brief details of search results are presented as a table in the ‘Result’ page, while more detailed descriptions such as PubMed ID and description of the reference are displayed in the ‘Detail’ page reached by selecting ‘more’. When selecting the specific RNA/Protein in the ‘Detail’ page (Supplementary Figure S2), the summary page presents more associated information of ncRNA or protein, such as official descriptions, functions (Gene ontology), pathways (KEGG and Biocarta), diseases (OMIM and Genetic Association Database) and interactors (HPRD and BIOGRID).

The ‘Browse page’ enables users to browse the database in four different ways: ‘Browse by Virus’, ‘Browse by Host Species’, ‘Browse by ncRNA Category’ and ‘Browse by ICTV’. All of the virus–host ncRNA-associated interaction information is presented by selecting on each entry.

VISUALIZATION

To help users interactively analyze virus–host ncRNA-associated interactions online, ViRBase provides a visualization function by embedding the Cytoscape web tool (<http://cytoscapeweb.cytoscape.org/>) (15) to highlight/visualize the interaction network or sub-network of a principal interaction. As the compelling visualization architecture is pan-and-zoom, in the ‘Visualization’ page (Figure 2) users can obtain a global view of the virus–host ncRNA-associated interaction network for a specific virus and host. At the bottom of the ‘Visualization’ page, ViRBase also provides the potential Gene Ontology functions and KEGG pathways by performing enrichment analysis of these virus–host ncRNA-associated interaction genes for specific virus and host, including HIV, HBV, H1N1, HCV, etc.

In the visualization option of the ‘Network’ page (Figure 2), virus–host ncRNA-associated interaction sub-networks can also be rapidly and independently represented for specific ncRNA-associated interactions by embedding interactive networks with Cytoscape Web. The ‘First Node’ or ‘Second Node’ option represents the sub-network of interacting RNA/Protein with the first or second interaction RNA/Protein. The ‘Both the Nodes’ option represents the sub-network of interacting RNA/Protein with both interaction nodes. The ‘First Neighbor’ represents the sub-network of direct interaction with the seed node. The ‘Second Neighbor’ represents the sub-network of direct and second-step interactions with the seed node. Interaction sub-networks based on the two nodes of this in-

teraction may help researchers to represent all interacting partners immediately. Thus, multiple RNA/Protein data resources can be combined in a single visualization for each RNA/Protein and its interaction partner. Users can observe specific RNA/Protein pairs within the virus–host ncRNA-associated interaction network and the ‘Selection of the Layout’ option can provide various layout types for this sub-network.

PREDICTED BINDING SITES

The identification of RNA-associated binding sites can provide valuable insights into the underlying regulatory mechanisms of various ncRNAs, thus the ‘Binding’ page of ViRBase also incorporates several useful tools to analyze the predicted binding site information for these interactions (Figure 3). For ncRNA–RNA interactions, the binding sites and scores are predicted according to miRanda (16) and Rsearch (17). For ncRNA–protein interactions, RNA-binding residues and scores are computed by bindN+ (18) and RNAbindR (19), while the experimentally verified RNA-binding sites documented in RBPBD (20) are also integrated. The parameters used by these tools were documented in the ‘Parameter’ field of the ‘Help’ page.

DISCUSSION AND FUTURE DIRECTIONS

Several viral databases centered on proteins, such as Virus-MINT (21), VirHostNet (22), HIV-1 (23) and the PHISTO (24) database, have been recently constructed and have led to a more comprehensive understanding of viral and host protein function in viral infection. However, viral protein–protein interactions are perhaps only half of the story in viral infection, as increasing numbers of ncRNAs with active involvement in viral infection are reported. Complemented by these useful databases, we developed the ViRBase database by manually collating ncRNA-associated interactions in viral infection. Thus, ViRBase provides an overview of the ncRNA-associated interaction network between viruses and hosts. To our knowledge, this is the first database focusing on ncRNA-associated interactions in viral infection. We hope this resource will bridge the gap in ncRNAs and viral research, and stimulate further interest in exploring the role of diverse ncRNAs in the virus–host network. ViRBase will be of particular interest to the viral science community and will facilitate virologists in unraveling the role of diverse ncRNAs in viral infection. In the future, we also plan to integrate experimentally verified viral protein interactions, providing a more useful resource for a better understanding of the functional organization in viral infection. Finally, we will continue to collate ncRNA-associated interaction reference data and update ViRBase.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National High Technology Research and Development Program of China [2014AA021102]; Major State Basic Research Development Program of China [2014CB910504];

National Natural Science Foundation of China [31100901, 81300253, 91129710, 61170154]; China Postdoctoral Science Foundation [2013M531064, 2014T70363]; Natural Science Foundation of Heilongjiang Province of China [QC2010012, ZD201114]; Heilongjiang Postdoctoral Foundation [LBH-Z12171]; Scientific Research Fund of Heilongjiang Provincial Education Department [12541426]. Funding for open access charge: National High Technology Research and Development Program of China [2014AA021102]; Major State Basic Research Development Program of China [2014CB910504]; National Natural Science Foundation of China [31100901, 81300253, 91129710, 61170154]; China Postdoctoral Science Foundation [2013M531064, 2014T70363]; Natural Science Foundation of Heilongjiang Province of China [QC2010012, ZD201114]; Heilongjiang Postdoctoral Foundation [LBH-Z12171]; Scientific Research Fund of Heilongjiang Provincial Education Department [12541426].

Conflict of interest statement. None declared.

REFERENCES

1. Saayman, S., Ackley, A., Turner, A.M., Famiglietti, M., Bosque, A., Clemson, M., Planelles, V. and Morris, K.V. (2014) An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol. Ther.*, **22**, 1164–1175.
2. Skalsky, R.L. and Cullen, B.R. (2010) Viruses, microRNAs, and host interactions. *Annu. Rev. Microbiol.*, **64**, 123–141.
3. Gottwein, E. and Cullen, B.R. (2008) Viral and cellular microRNAs as determinants of viral pathogenesis and immunity. *Cell Host Microbe*, **3**, 375–387.
4. Mocarski, E.S., Upton, J.W. and Kaiser, W.J. (2012) Viral infection and the evolution of caspase 8-regulated apoptotic and necrotic death pathways. *Nat. Rev. Immunol.*, **12**, 79–88.
5. Ghosh, Z., Mallick, B. and Chakrabarti, J. (2009) Cellular versus viral microRNAs in host-virus interaction. *Nucleic Acids Res.*, **37**, 1035–1048.
6. Klase, Z., Houzet, L. and Jeang, K.T. (2012) MicroRNAs and HIV-1: complex interactions. *J. Biol. Chem.*, **287**, 40884–40890.
7. Li, Y., Chan, E.Y., Li, J., Ni, C., Peng, X., Rosenzweig, E., Tumpey, T.M. and Katze, M.G. (2010) MicroRNA expression and virulence in pandemic influenza virus-infected mice. *J. Virol.*, **84**, 3023–3032.
8. Peng, X., Gralinski, L., Armour, C.D., Ferris, M.T., Thomas, M.J., Prohl, S., Bradel-Tretheway, B.G., Korth, M.J., Castle, J.C., Biery, M.C. *et al.* (2010) Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *mBio*, **1**, e00206–e00210.
9. Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E. and Mattick, J.S. (2011) lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res.*, **39**, D146–D151.
10. Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
11. Xie, J., Zhang, M., Zhou, T., Hua, X., Tang, L. and Wu, W. (2007) Sno/scaRNABase: a curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Res.*, **35**, D183–D187.
12. Lestrade, L. and Weber, M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
13. King, A.M.Q., Adams, M.J., Carstens, E.B. and Lefkowitz, E.J. (2012) *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press.
14. Fauquet, C.M. and Fargette, D. (2005) International committee on taxonomy of viruses and the 3,142 unassigned species. *Virol. J.*, **2**, 64.
15. Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q. and Bader, G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
16. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.
17. Wenzel, A., Akbasli, E. and Gorodkin, J. (2012) RIssearch: fast RNA-RNA interaction search using a simplified nearest-neighbor energy model. *Bioinformatics*, **28**, 2738–2746.
18. Wang, L., Huang, C., Yang, M.Q. and Yang, J.Y. (2010) BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Syst. Biol.*, **4**(Suppl. 1), S3.
19. Terribilini, M., Sander, J.D., Lee, J.H., Zaback, P., Jernigan, R.L., Honavar, V. and Dobbs, D. (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.*, **35**, W578–W584.
20. Cook, K.B., Kazan, H., Zuberi, K., Morris, Q. and Hughes, T.R. (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, **39**, D301–D308.
21. Chatr-aryamontri, A., Ceol, A., Peluso, D., Nardozza, A., Panni, S., Sacco, F., Tinti, M., Smolyar, A., Castagnoli, L., Vidal, M. *et al.* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res.*, **37**, D669–D673.
22. Navratil, V., de Chasse, B., Meyniel, L., Delmotte, S., Gautier, C., Andre, P., Lotteau, V. and Rabourdin-Combe, C. (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.*, **37**, D661–D668.
23. Fu, W., Sanders-Beer, B.E., Katz, K.S., Maglott, D.R., Pruitt, K.D. and Ptak, R.G. (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res.*, **37**, D417–D422.
24. Durmus Tekir, S., Cakir, T., Ardic, E., Sayilirbas, A.S., Konuk, G., Konuk, M., Sariyer, H., Ugurlu, A., Karadeniz, I., Ozgur, A. *et al.* (2013) PHISTO: pathogen-host interaction search tool. *Bioinformatics*, **29**, 1357–1358.